

CLAIMS

1 1. A method of processing a natural language
2 input, comprising:

3 identifying N-best part-of-speech (POS) sequences
4 corresponding to the natural language input;
5 identifying a likely base noun phrase (baseNP)
6 sequence based on the N-best POS sequences
7 identified; and
8 outputting the likely baseNP sequence.

9 2. The method of claim 1 wherein identifying a
10 likely baseNP sequence comprises:

11 identifying a plurality of baseNP sequences for
12 each of the N-best POS tag sequences; and
13 calculating which of the plurality of baseNP
14 sequences is most likely.

15 3. The method of claim 2 wherein calculating
16 which of the plurality of baseNP sequences is most
17 likely comprises:

18 calculating a likely baseNP sequence that is most
19 likely based on lexical information indicative
20 of a position of words in the natural language
21 input relative to baseNPs identified in the
22 baseNP sequences.

23 4. The method of claim 3 wherein calculating a
24 likely baseNP sequence that is most likely based on
25 lexical information comprises:

calculating a likely baseNP sequence that is most likely based on lexical information indicative of POS tags assigned to the words in the natural language input.

5. The method of claim 3 wherein calculating a likely baseNP sequence comprises:

calculating a likely baseNP sequence based on the lexical information for every word in the natural language input.

6. The method of claim 2 wherein the natural language input comprises a sentence and wherein calculating which of the plurality of baseNP sequences is most likely comprises:

calculating which of the plurality of baseNP sequences is most likely over the entire sentence.

7. The method of claim 3 wherein a baseNP rule comprises a sequence of POS tags corresponding to words in the natural language input identified as a baseNP and wherein calculating a likely baseNP sequence comprises:

calculating a probability of POS tags and baseNP rules, given their context.

8. The method of claim 7 wherein calculating a probability of POS tags and baseNP rules comprises:

calculating the probability of POS tags and baseNP rules given n prior POS tags or baseNP rules.

9. The method of claim 1 wherein identifying a likely baseNP sequence includes:

calculating a probability of each of the N-best POS sequences given the natural language input.

10. A natural language processing system for processing a natural language input, comprising:

a base noun phrase (baseNP) identifier configured to receive N-best part-of-speech (POS) tag sequences for the natural language input and identify a likely baseNP sequence of baseNPs corresponding to the natural language input, given the N-best POS tag sequences.

11. The system of claim 10 and further comprising: a POS tagger, coupled to the baseNP identifier, receiving the natural language input and calculating the N-best POS tag sequences corresponding to the natural language input.

12. The system of claim 11 wherein the baseNP identifier is configured to identify a plurality of baseNP sequences for each of the POS tag sequences and calculate which of the plurality of baseNP sequences is the likely baseNP sequence.

1 13. The system of claim 12 wherein the baseNP
2 identifier further comprises:

3 a unified statistical model that includes lexical
4 information indicative of a position of words
5 in the natural language input relative to
6 baseNPs identified in the baseNP sequences.

7
8 14. The system of claim 13 wherein a baseNP rule comprises
9 a sequence of POS tags corresponding to words in the
10 natural language input that are identified as a baseNP and
11 wherein the unified statistical model includes a baseNP
12 rule component for calculating a probability of POS tags
13 and baseNP rules, given contextual information.

14 15. The system of claim 14 wherein the baseNP rule
15 component is configured to calculate the probability of POS
16 tags and baseNP rules, given n prior POS tags and baseNP
17 rules.

18 16. The system of claim 15 wherein the natural
19 language input comprises a sentence and wherein the
20 unified statistical model is configured for calculating
21 which of the plurality of baseNP sequences is most
22 likely over the entire sentence.

23 17. A method of processing a linguistic input, comprising:
24 identifying N-best part-of-speech (POS) sequences
25 corresponding to the linguistic input;

identifying one or more base noun phrases (baseNPs)
for each of the N-best POS sequences to form a
plurality of different possible baseNP sequences
corresponding to each of the POS sequences; for
each baseNP sequence, identifying whether it is a
likely baseNP sequence based on a probability of
the associated POS sequence and a probability of
the baseNP sequence, given lexical information
indicative of a position of words in the
linguistic input relative to the baseNPs
identified in the baseNP sequence; and
outputting the likely baseNP sequence identified.

18. The method of claim 17 wherein identifying one or more
baseNPs for each of the N-best POS sequences comprises:

identifying baseNP rules for each of the N-best POS
sequences, the baseNP rules comprising a
sequence of POS tags corresponding to words in
the linguistic input identified as a baseNP.

19. The method of claim 18 wherein identifying whether
each baseNP sequence is a likely baseNP sequence,
comprises:

calculating a probability of POS tags and baseNP
rules, given n prior POS tags or baseNP rules
in the POS sequence.

20. The method of claim 19 wherein the linguistic input
comprises a sentence and wherein identifying whether each
baseNP sequence is a likely baseNP sequence comprises:

identifying whether each baseNP sequence is a
likely baseNP sequence over the entire
sentence.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

092733E-060404